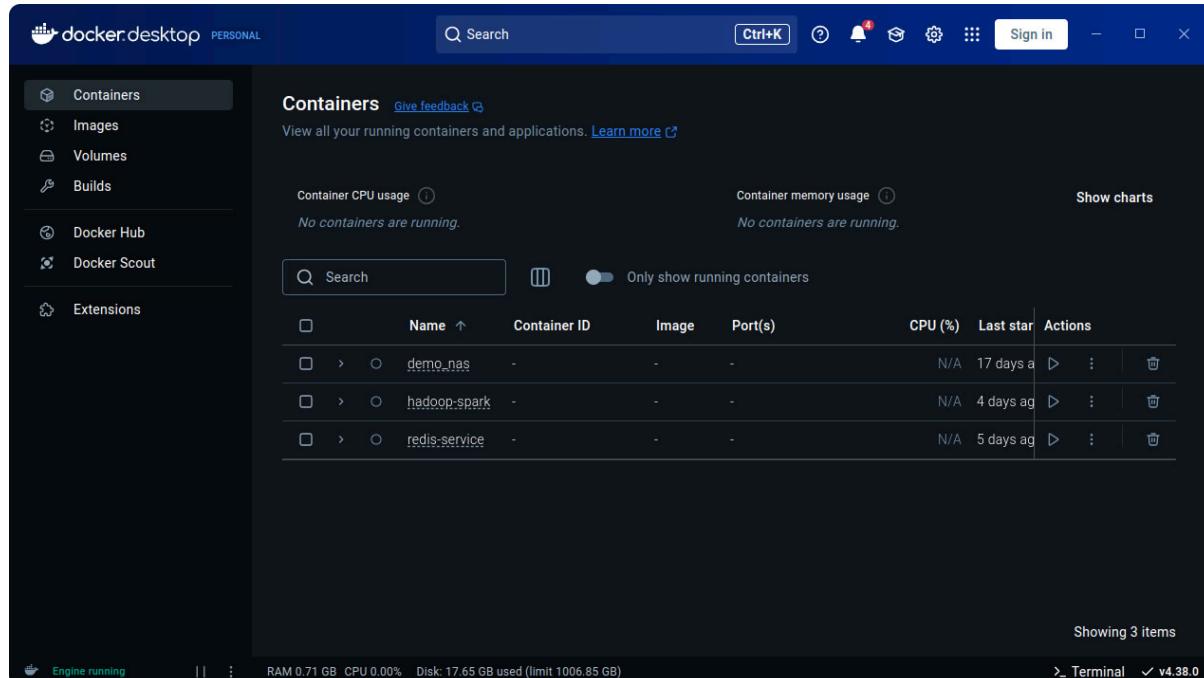


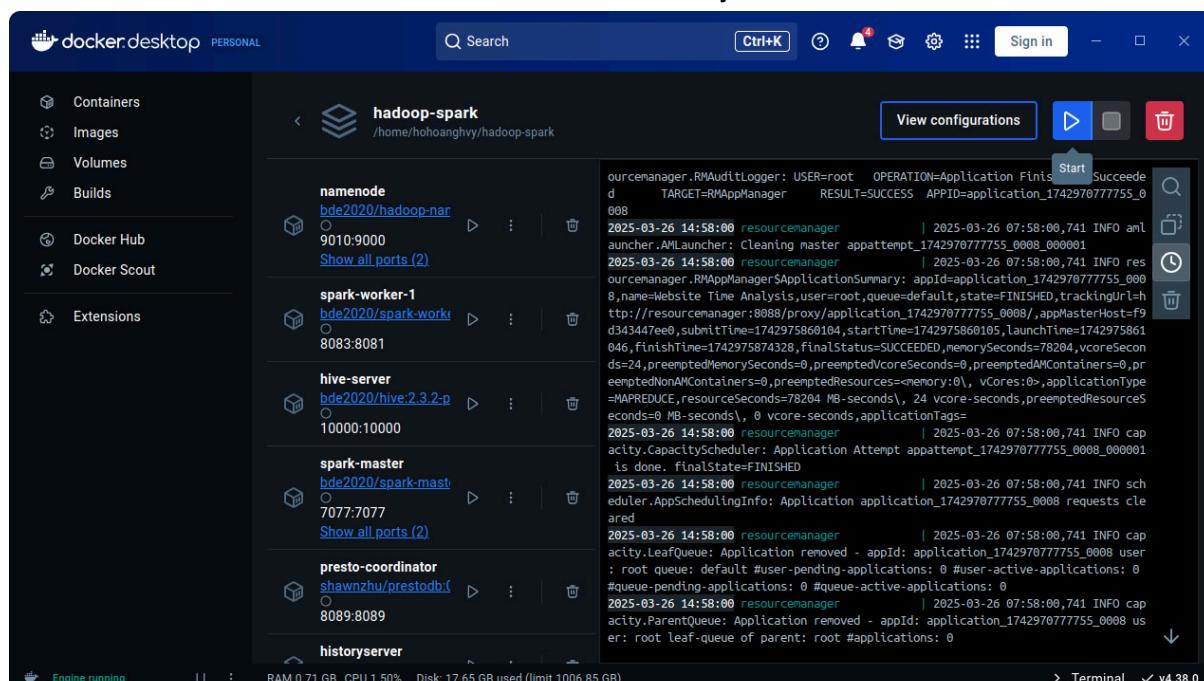
Bài 1: Giới thiệu từng bước thực hiện và minh họa bằng ảnh chụp màn hình. Cấu trúc cây thư mục

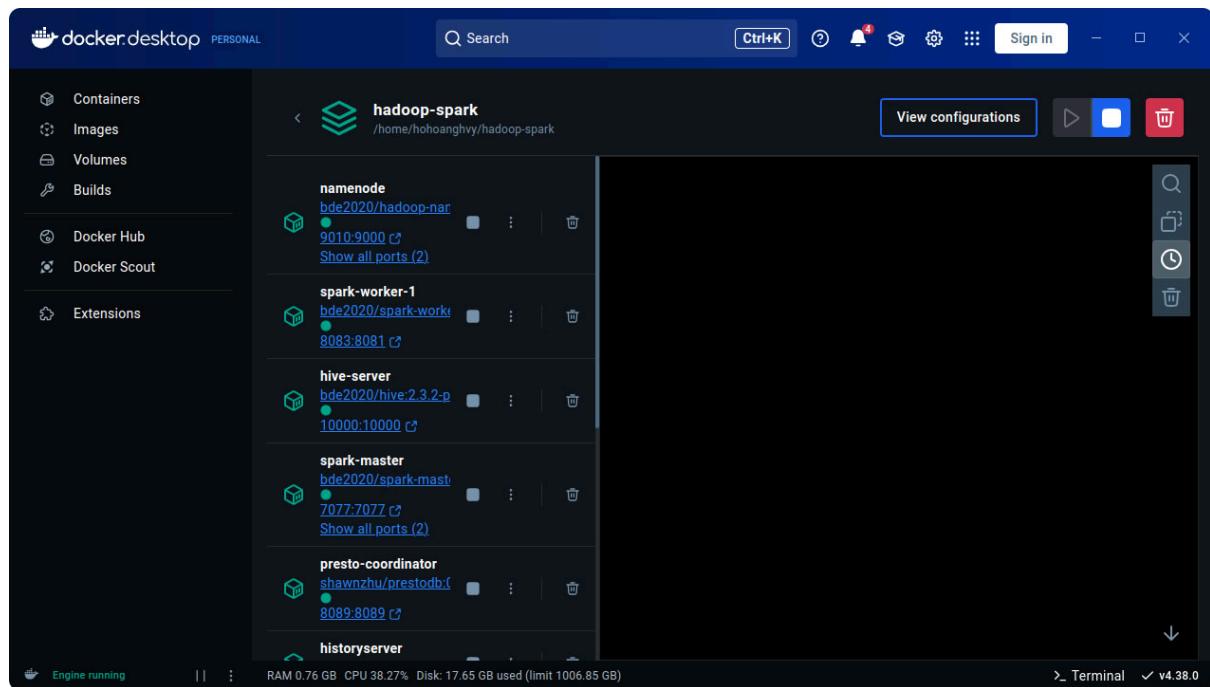
- /PhanLop/NaiveBayes
- /PhanLop/DecisionTree
- /PhanLop/SVM
- /GomCum/kMeans
- /GomCum/DBSCAN

B1: Open docker desktop and choose docker container for hadoop. In my computer its name is hadoop-spark



B2: After that run this container and wait for it until every container inside turn blue





B3: Run docker exec -it namenode bash in terminal to enter docker container bash to run command

A screenshot of a terminal window. The title bar shows the user's name 'hohoanghvy@hvy-dev:~'. The command 'docker exec -it namenode bash' is typed into the terminal. The terminal window has a dark background and light-colored text.

```
docker exec -it namenode bash
root@bbd0b037ecb0:/# |
```

B4: Run hdfs command to create our expected folder structure with -mkdir then use -ls command to check whether the creation is successful.

```
docker exec -it namenode bash
root@bbd0b037ecb0:/# hdfs dfs -mkdir -p /PhanLop/NaiveBayes /PhanLop/DecisionTree
/PhanLop/SVM /GomCum/kMeans /GomCum/DBSCAN
root@bbd0b037ecb0:/# hdfs dfs -ls
ls: `.': No such file or directory
root@bbd0b037ecb0:/# hdfs dfs -ls /
Found 6 items
drwxr-xr-x  - root supergroup          0 2025-03-30 04:50 /GomCum
drwxr-xr-x  - root supergroup          0 2025-03-30 04:50 /PhanLop
drwxrwxrwt  - root root              0 2025-03-25 16:53 /app-logs
drwxr-xr-x  - root supergroup          0 2025-03-25 15:58 /rmstate
drwxrwxr-x  - root supergroup          0 2025-03-25 16:53 /tmp
drwxr-xr-x  - root supergroup          0 2025-03-25 16:49 /user
root@bbd0b037ecb0:/#
```

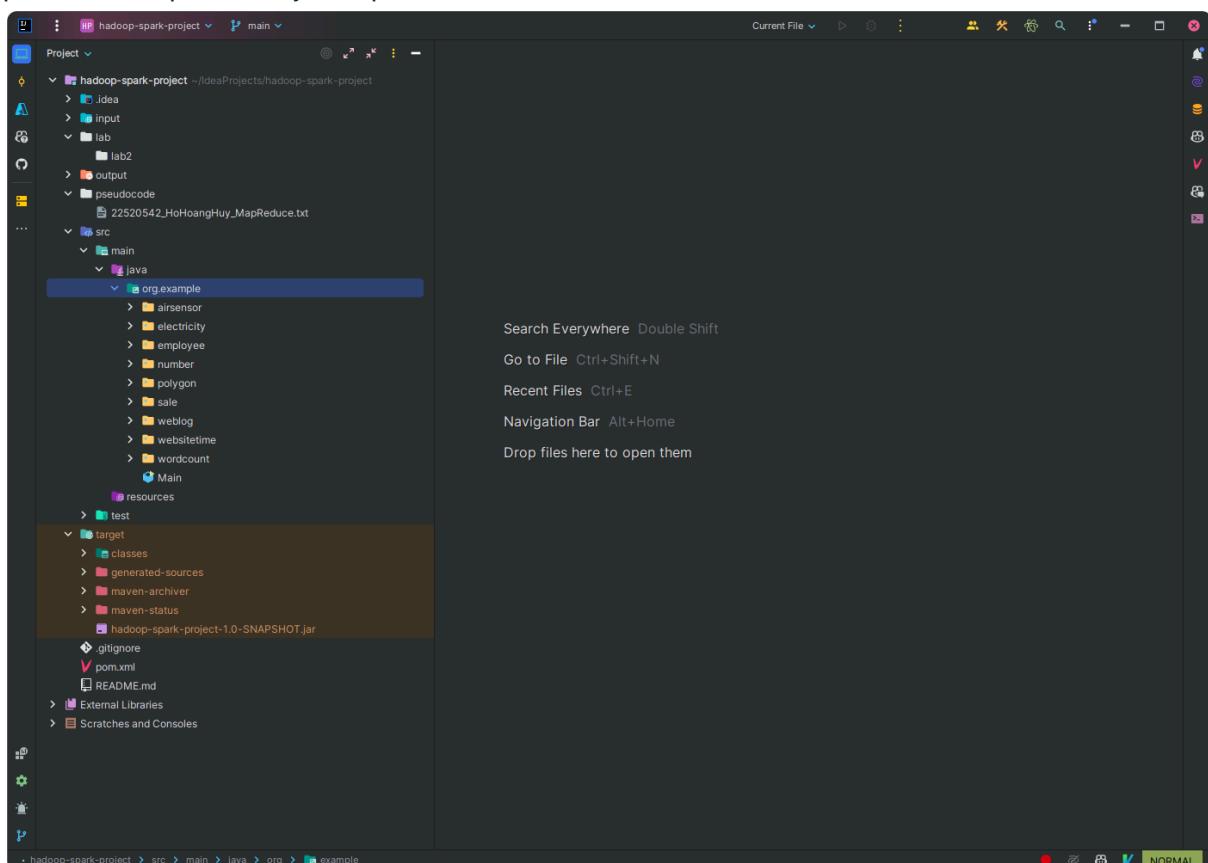
```

root@bbd0b037ecb0:/# hdfs dfs -mkdir -p /PhanLop/NaiveBayes /PhanLop/DecisionTree
/PhanLop/SVM /GomCum/kMeans /GomCum/DBSCAN
root@bbd0b037ecb0:/# hdfs dfs -ls
ls: `.' : No such file or directory
root@bbd0b037ecb0:/# hdfs dfs -ls /
Found 6 items
drwxr-xr-x  - root supergroup          0 2025-03-30 04:50 /GomCum
drwxr-xr-x  - root supergroup          0 2025-03-30 04:50 /PhanLop
drwxrwxrwt - root root                0 2025-03-25 16:53 /app-logs
drwxr-xr-x  - root supergroup          0 2025-03-25 15:58 /rmstate
drwxrwxr-x  - root supergroup          0 2025-03-25 16:53 /tmp
drwxr-xr-x  - root supergroup          0 2025-03-25 16:49 /user
root@bbd0b037ecb0:/# hdfs dfs -ls /GomCum
Found 2 items
drwxr-xr-x  - root supergroup          0 2025-03-30 04:50 /GomCum/DBSCAN
drwxr-xr-x  - root supergroup          0 2025-03-30 04:50 /GomCum/kMeans
root@bbd0b037ecb0:/# hdfs dfs -ls /PhanLop
Found 3 items
drwxr-xr-x  - root supergroup          0 2025-03-30 04:50 /PhanLop/DecisionTree
drwxr-xr-x  - root supergroup          0 2025-03-30 04:50 /PhanLop/NaiveBayes
drwxr-xr-x  - root supergroup          0 2025-03-30 04:50 /PhanLop/SVM
root@bbd0b037ecb0:#

```

Bài 2: Nộp minh họa các bước làm WordCount v2.0

B1: Open Intelij and then create the wordcountv2 package and also create a class WordCount 2 with a piece of code provided by the apache



The screenshot shows the IntelliJ IDEA interface with the project 'hadoop-spark-project' open. The left sidebar displays the project structure, including packages like lab, output, pseudocode, and src/main/java/org/example. The right pane shows the code editor for WordCount2.java. The code implements a Mapper that tokenizes input text and counts word occurrences. It includes imports for Java util and Apache Hadoop classes, and defines a TokenizerMapper class that extends Mapper<Object, Text, Text, IntWritable>. The setup method handles configuration and file reading, while the map method processes each line of input to produce word counts.

```

import java.io.IOException;
import java.net.URL;
import java.util.ArrayList;
import java.util.HashSet;
import java.util.List;
import java.util.Set;
import java.util.StringTokenizer;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.Counter;
import org.apache.hadoop.util.GenericOptionsParser;
import org.apache.hadoop.util.StringUtils;

```

```

public class WordCount2 { new * }

public static class TokenizerMapper extends Mapper<Object, Text, Text, IntWritable> {
    static enum CountersEnum { INPUT_WORDS } 2 usages new *
    private final static IntWritable one = new IntWritable(1); 1 usage
    private Text word = new Text(); 2 usages

    private boolean caseSensitive; 2 usages
    private Set<String> patternsToSkip = new HashSet<String>(); 2 usages

    private Configuration conf; 4 usages
    private BufferedReader fis; 2 usages

    @Override new *
    public void setup(Context context) throws IOException,
        InterruptedException {
        conf = context.getConfiguration();
        caseSensitive = conf.getBoolean("wordcount.case.sensitive", defaultValue: true);
        if (conf.getBoolean("wordcount.skip.patterns", defaultValue: false)) {
            URI[] patternsURIs = Job.getInstance(conf).getCacheFiles();
            for (URI patternsURI : patternsURIs) {
                Path patternsPath = new Path(patternsURI.getPath());
            }
        }
    }
}

```

B2: Run mvn clean package to build the jar file into our target folder which persist which a /user/hadoop/target in hadoop container

The screenshot shows the IntelliJ IDEA interface with the project 'hadoop-spark-project' open. The left sidebar displays the project structure. The right pane shows the terminal output of the 'mvn clean package' command. The output shows Maven scanning for projects, building the 'hadoop-spark-project' version 1.0-SNAPSHOT, and creating a JAR file in the target directory. It also shows the compilation of source files, the execution of unit tests using Maven Surefire, and the creation of a distribution JAR file named 'hadoop-spark-project-1.0-SNAPSHOT.jar'.

```

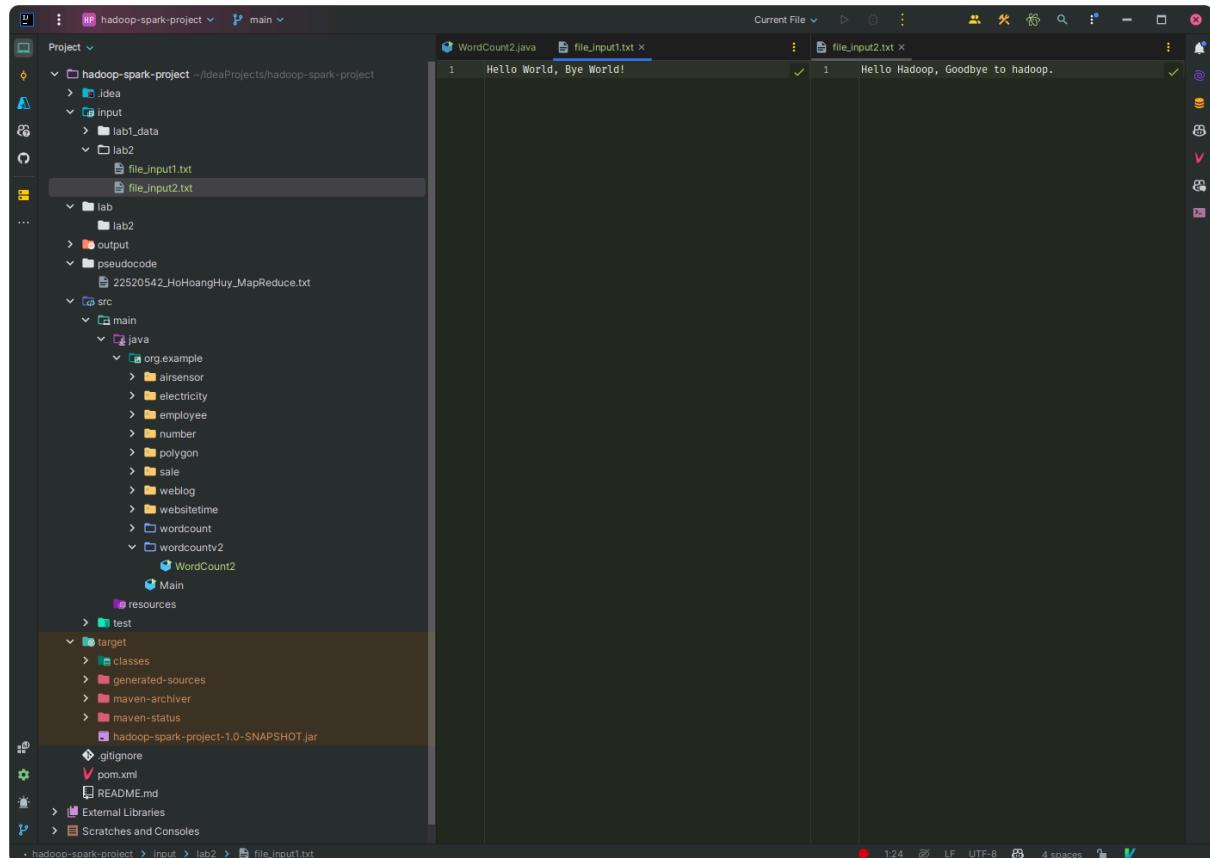
16 import 'hadoop-spark-project' git:(main) mvn clean package
17 [INFO] Scanning for projects...
18 [INFO] ------------------------------------------------------------------------
19 [INFO] -----> org.example:hadoop-spark-project <-----
20 [INFO] Building hadoop-spark-project 1.0-SNAPSHOT
21 [INFO] [jar]
22 [INFO] --- maven-clean-plugin:2.5:clean (default-clean) @ hadoop-spark-project ---
23 [INFO] Deleting /home/hoahoanghvy/IdeaProjects/hadoop-spark-project/target
24 [INFO] --- maven-resources-plugin:2.6:resources (default-resources) @ hadoop-spark-project
25 [INFO] --- maven-compiler-plugin:3.1:compile (default-compile) @ hadoop-spark-project
26 [INFO] Changes detected - recompiling the module!
27 [INFO] Compiling 32 source files to /home/hoahoanghvy/IdeaProjects/hadoop-spark-project/target/classes
28 [INFO] 
29 [INFO] Using 'UTF-8' encoding to copy filtered resources.
30 [INFO] Copying 0 resource
31 [INFO] 
32 [INFO] --- maven-compiler-plugin:3.1:compile (default-compile) @ hadoop-spark-project
33 [INFO] Changes detected - recompiling the module!
34 [INFO] Compiling 32 source files to /home/hoahoanghvy/IdeaProjects/hadoop-spark-project/target/classes
35 [INFO] 
36 [INFO] --- maven-resources-plugin:2.6:testResources (default-testResources) @ hadoop-spark-project
37 [INFO] skip non existing resourceDirectory /home/hoahoanghvy/IdeaProjects/hadoop-spark-project/src/test/resources
38 [INFO] 
39 [INFO] --- maven-compiler-plugin:3.1:testCompile (default-testCompile) @ hadoop-spark-project
40 [INFO] 
41 [INFO] Nothing to compile - all classes are up to date
42 [INFO] 
43 [INFO] --- maven-surefire-plugin:2.12.4:test (default-test) @ hadoop-spark-project
44 [INFO] No tests to run.
45 [INFO] 
46 [INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ hadoop-spark-project
47 [INFO] Building jar: /home/hoahoanghvy/IdeaProjects/hadoop-spark-project/target/hadoop-spark-project-1.0-SNAPSHOT.jar
48 [INFO] 
49 [INFO] BUILD SUCCESS
50 [INFO] 
51 [INFO] Total time: 2.002 s
52 [INFO] Finished at: 2025-03-30T12:00:41+07:00
53 [INFO] 
54 [INFO] 
55 [INFO] 
56 [INFO] 
57 [INFO] 
58 [INFO] 
59 [INFO] 
60 [INFO] 
61 [INFO] 
62 [INFO] 
63 [INFO] 
64 [INFO] 
65 [INFO] 
66 [INFO] 
67 [INFO] 
68 [INFO] 
69 [INFO] 
70 [INFO] 
71 [INFO] 
72 [INFO] 
73 [INFO] 
74 [INFO] 
75 [INFO] 
76 [INFO] 
77 [INFO] 
78 [INFO] 
79 [INFO] 
80 [INFO] 
81 [INFO] 
82 [INFO] 
83 [INFO] 
84 [INFO] 
85 [INFO] 
86 [INFO] 
87 [INFO] 
88 [INFO] 
89 [INFO] 
90 [INFO] 
91 [INFO] 
92 [INFO] 
93 [INFO] 
94 [INFO] 
95 [INFO] 
96 [INFO] 
97 [INFO] 
98 [INFO] 
99 [INFO] 
100 [INFO] 
101 [INFO] 
102 [INFO] 
103 [INFO] 
104 [INFO] 
105 [INFO] 
106 [INFO] 
107 [INFO] 
108 [INFO] 
109 [INFO] 
110 [INFO] 
111 [INFO] 
112 [INFO] 
113 [INFO] 
114 [INFO] 
115 [INFO] 
116 [INFO] 
117 [INFO] 
118 [INFO] 
119 [INFO] 
120 [INFO] 
121 [INFO] 
122 [INFO] 
123 [INFO] 
124 [INFO] 
125 [INFO] 
126 [INFO] 
127 [INFO] 
128 [INFO] 
129 [INFO] 
130 [INFO] 
131 [INFO] 
132 [INFO] 
133 [INFO] 
134 [INFO] 
135 [INFO] 
136 [INFO] 
137 [INFO] 
138 [INFO] 
139 [INFO] 
140 [INFO] 
141 [INFO] 
142 [INFO] 
143 [INFO] 
144 [INFO] 
145 [INFO] 
146 [INFO] 
147 [INFO] 
148 [INFO] 
149 [INFO] 
150 [INFO] 
151 [INFO] 
152 [INFO] 
153 [INFO] 
154 [INFO] 
155 [INFO] 
156 [INFO] 
157 [INFO] 
158 [INFO] 
159 [INFO] 
160 [INFO] 
161 [INFO] 
162 [INFO] 
163 [INFO] 
164 [INFO] 
165 [INFO] 
166 [INFO] 
167 [INFO] 
168 [INFO] 
169 [INFO] 
170 [INFO] 
171 [INFO] 
172 [INFO] 
173 [INFO] 
174 [INFO] 
175 [INFO] 
176 [INFO] 
177 [INFO] 
178 [INFO] 
179 [INFO] 
180 [INFO] 
181 [INFO] 
182 [INFO] 
183 [INFO] 
184 [INFO] 
185 [INFO] 
186 [INFO] 
187 [INFO] 
188 [INFO] 
189 [INFO] 
190 [INFO] 
191 [INFO] 
192 [INFO] 
193 [INFO] 
194 [INFO] 
195 [INFO] 
196 [INFO] 
197 [INFO] 
198 [INFO] 
199 [INFO] 
200 [INFO] 
201 [INFO] 
202 [INFO] 
203 [INFO] 
204 [INFO] 
205 [INFO] 
206 [INFO] 
207 [INFO] 
208 [INFO] 
209 [INFO] 
210 [INFO] 
211 [INFO] 
212 [INFO] 
213 [INFO] 
214 [INFO] 
215 [INFO] 
216 [INFO] 
217 [INFO] 
218 [INFO] 
219 [INFO] 
220 [INFO] 
221 [INFO] 
222 [INFO] 
223 [INFO] 
224 [INFO] 
225 [INFO] 
226 [INFO] 
227 [INFO] 
228 [INFO] 
229 [INFO] 
230 [INFO] 
231 [INFO] 
232 [INFO] 
233 [INFO] 
234 [INFO] 
235 [INFO] 
236 [INFO] 
237 [INFO] 
238 [INFO] 
239 [INFO] 
240 [INFO] 
241 [INFO] 
242 [INFO] 
243 [INFO] 
244 [INFO] 
245 [INFO] 
246 [INFO] 
247 [INFO] 
248 [INFO] 
249 [INFO] 
250 [INFO] 
251 [INFO] 
252 [INFO] 
253 [INFO] 
254 [INFO] 
255 [INFO] 
256 [INFO] 
257 [INFO] 
258 [INFO] 
259 [INFO] 
260 [INFO] 
261 [INFO] 
262 [INFO] 
263 [INFO] 
264 [INFO] 
265 [INFO] 
266 [INFO] 
267 [INFO] 
268 [INFO] 
269 [INFO] 
270 [INFO] 
271 [INFO] 
272 [INFO] 
273 [INFO] 
274 [INFO] 
275 [INFO] 
276 [INFO] 
277 [INFO] 
278 [INFO] 
279 [INFO] 
280 [INFO] 
281 [INFO] 
282 [INFO] 
283 [INFO] 
284 [INFO] 
285 [INFO] 
286 [INFO] 
287 [INFO] 
288 [INFO] 
289 [INFO] 
290 [INFO] 
291 [INFO] 
292 [INFO] 
293 [INFO] 
294 [INFO] 
295 [INFO] 
296 [INFO] 
297 [INFO] 
298 [INFO] 
299 [INFO] 
300 [INFO] 
301 [INFO] 
302 [INFO] 
303 [INFO] 
304 [INFO] 
305 [INFO] 
306 [INFO] 
307 [INFO] 
308 [INFO] 
309 [INFO] 
310 [INFO] 
311 [INFO] 
312 [INFO] 
313 [INFO] 
314 [INFO] 
315 [INFO] 
316 [INFO] 
317 [INFO] 
318 [INFO] 
319 [INFO] 
320 [INFO] 
321 [INFO] 
322 [INFO] 
323 [INFO] 
324 [INFO] 
325 [INFO] 
326 [INFO] 
327 [INFO] 
328 [INFO] 
329 [INFO] 
330 [INFO] 
331 [INFO] 
332 [INFO] 
333 [INFO] 
334 [INFO] 
335 [INFO] 
336 [INFO] 
337 [INFO] 
338 [INFO] 
339 [INFO] 
340 [INFO] 
341 [INFO] 
342 [INFO] 
343 [INFO] 
344 [INFO] 
345 [INFO] 
346 [INFO] 
347 [INFO] 
348 [INFO] 
349 [INFO] 
350 [INFO] 
351 [INFO] 
352 [INFO] 
353 [INFO] 
354 [INFO] 
355 [INFO] 
356 [INFO] 
357 [INFO] 
358 [INFO] 
359 [INFO] 
360 [INFO] 
361 [INFO] 
362 [INFO] 
363 [INFO] 
364 [INFO] 
365 [INFO] 
366 [INFO] 
367 [INFO] 
368 [INFO] 
369 [INFO] 
370 [INFO] 
371 [INFO] 
372 [INFO] 
373 [INFO] 
374 [INFO] 
375 [INFO] 
376 [INFO] 
377 [INFO] 
378 [INFO] 
379 [INFO] 
380 [INFO] 
381 [INFO] 
382 [INFO] 
383 [INFO] 
384 [INFO] 
385 [INFO] 
386 [INFO] 
387 [INFO] 
388 [INFO] 
389 [INFO] 
390 [INFO] 
391 [INFO] 
392 [INFO] 
393 [INFO] 
394 [INFO] 
395 [INFO] 
396 [INFO] 
397 [INFO] 
398 [INFO] 
399 [INFO] 
400 [INFO] 
401 [INFO] 
402 [INFO] 
403 [INFO] 
404 [INFO] 
405 [INFO] 
406 [INFO] 
407 [INFO] 
408 [INFO] 
409 [INFO] 
410 [INFO] 
411 [INFO] 
412 [INFO] 
413 [INFO] 
414 [INFO] 
415 [INFO] 
416 [INFO] 
417 [INFO] 
418 [INFO] 
419 [INFO] 
420 [INFO] 
421 [INFO] 
422 [INFO] 
423 [INFO] 
424 [INFO] 
425 [INFO] 
426 [INFO] 
427 [INFO] 
428 [INFO] 
429 [INFO] 
430 [INFO] 
431 [INFO] 
432 [INFO] 
433 [INFO] 
434 [INFO] 
435 [INFO] 
436 [INFO] 
437 [INFO] 
438 [INFO] 
439 [INFO] 
440 [INFO] 
441 [INFO] 
442 [INFO] 
443 [INFO] 
444 [INFO] 
445 [INFO] 
446 [INFO] 
447 [INFO] 
448 [INFO] 
449 [INFO] 
450 [INFO] 
451 [INFO] 
452 [INFO] 
453 [INFO] 
454 [INFO] 
455 [INFO] 
456 [INFO] 
457 [INFO] 
458 [INFO] 
459 [INFO] 
460 [INFO] 
461 [INFO] 
462 [INFO] 
463 [INFO] 
464 [INFO] 
465 [INFO] 
466 [INFO] 
467 [INFO] 
468 [INFO] 
469 [INFO] 
470 [INFO] 
471 [INFO] 
472 [INFO] 
473 [INFO] 
474 [INFO] 
475 [INFO] 
476 [INFO] 
477 [INFO] 
478 [INFO] 
479 [INFO] 
480 [INFO] 
481 [INFO] 
482 [INFO] 
483 [INFO] 
484 [INFO] 
485 [INFO] 
486 [INFO] 
487 [INFO] 
488 [INFO] 
489 [INFO] 
490 [INFO] 
491 [INFO] 
492 [INFO] 
493 [INFO] 
494 [INFO] 
495 [INFO] 
496 [INFO] 
497 [INFO] 
498 [INFO] 
499 [INFO] 
500 [INFO] 
501 [INFO] 
502 [INFO] 
503 [INFO] 
504 [INFO] 
505 [INFO] 
506 [INFO] 
507 [INFO] 
508 [INFO] 
509 [INFO] 
510 [INFO] 
511 [INFO] 
512 [INFO] 
513 [INFO] 
514 [INFO] 
515 [INFO] 
516 [INFO] 
517 [INFO] 
518 [INFO] 
519 [INFO] 
520 [INFO] 
521 [INFO] 
522 [INFO] 
523 [INFO] 
524 [INFO] 
525 [INFO] 
526 [INFO] 
527 [INFO] 
528 [INFO] 
529 [INFO] 
530 [INFO] 
531 [INFO] 
532 [INFO] 
533 [INFO] 
534 [INFO] 
535 [INFO] 
536 [INFO] 
537 [INFO] 
538 [INFO] 
539 [INFO] 
540 [INFO] 
541 [INFO] 
542 [INFO] 
543 [INFO] 
544 [INFO] 
545 [INFO] 
546 [INFO] 
547 [INFO] 
548 [INFO] 
549 [INFO] 
550 [INFO] 
551 [INFO] 
552 [INFO] 
553 [INFO] 
554 [INFO] 
555 [INFO] 
556 [INFO] 
557 [INFO] 
558 [INFO] 
559 [INFO] 
560 [INFO] 
561 [INFO] 
562 [INFO] 
563 [INFO] 
564 [INFO] 
565 [INFO] 
566 [INFO] 
567 [INFO] 
568 [INFO] 
569 [INFO] 
570 [INFO] 
571 [INFO] 
572 [INFO] 
573 [INFO] 
574 [INFO] 
575 [INFO] 
576 [INFO] 
577 [INFO] 
578 [INFO] 
579 [INFO] 
580 [INFO] 
581 [INFO] 
582 [INFO] 
583 [INFO] 
584 [INFO] 
585 [INFO] 
586 [INFO] 
587 [INFO] 
588 [INFO] 
589 [INFO] 
590 [INFO] 
591 [INFO] 
592 [INFO] 
593 [INFO] 
594 [INFO] 
595 [INFO] 
596 [INFO] 
597 [INFO] 
598 [INFO] 
599 [INFO] 
600 [INFO] 
601 [INFO] 
602 [INFO] 
603 [INFO] 
604 [INFO] 
605 [INFO] 
606 [INFO] 
607 [INFO] 
608 [INFO] 
609 [INFO] 
610 [INFO] 
611 [INFO] 
612 [INFO] 
613 [INFO] 
614 [INFO] 
615 [INFO] 
616 [INFO] 
617 [INFO] 
618 [INFO] 
619 [INFO] 
620 [INFO] 
621 [INFO] 
622 [INFO] 
623 [INFO] 
624 [INFO] 
625 [INFO] 
626 [INFO] 
627 [INFO] 
628 [INFO] 
629 [INFO] 
630 [INFO] 
631 [INFO] 
632 [INFO] 
633 [INFO] 
634 [INFO] 
635 [INFO] 
636 [INFO] 
637 [INFO] 
638 [INFO] 
639 [INFO] 
640 [INFO] 
641 [INFO] 
642 [INFO] 
643 [INFO] 
644 [INFO] 
645 [INFO] 
646 [INFO] 
647 [INFO] 
648 [INFO] 
649 [INFO] 
650 [INFO] 
651 [INFO] 
652 [INFO] 
653 [INFO] 
654 [INFO] 
655 [INFO] 
656 [INFO] 
657 [INFO] 
658 [INFO] 
659 [INFO] 
660 [INFO] 
661 [INFO] 
662 [INFO] 
663 [INFO] 
664 [INFO] 
665 [INFO] 
666 [INFO] 
667 [INFO] 
668 [INFO] 
669 [INFO] 
670 [INFO] 
671 [INFO] 
672 [INFO] 
673 [INFO] 
674 [INFO] 
675 [INFO] 
676 [INFO] 
677 [INFO] 
678 [INFO] 
679 [INFO] 
680 [INFO] 
681 [INFO] 
682 [INFO] 
683 [INFO] 
684 [INFO] 
685 [INFO] 
686 [INFO] 
687 [INFO] 
688 [INFO] 
689 [INFO] 
690 [INFO] 
691 [INFO] 
692 [INFO] 
693 [INFO] 
694 [INFO] 
695 [INFO] 
696 [INFO] 
697 [INFO] 
698 [INFO] 
699 [INFO] 
700 [INFO] 
701 [INFO] 
702 [INFO] 
703 [INFO] 
704 [INFO] 
705 [INFO] 
706 [INFO] 
707 [INFO] 
708 [INFO] 
709 [INFO] 
710 [INFO] 
711 [INFO] 
712 [INFO] 
713 [INFO] 
714 [INFO] 
715 [INFO] 
716 [INFO] 
717 [INFO] 
718 [INFO] 
719 [INFO] 
720 [INFO] 
721 [INFO] 
722 [INFO] 
723 [INFO] 
724 [INFO] 
725 [INFO] 
726 [INFO] 
727 [INFO] 
728 [INFO] 
729 [INFO] 
730 [INFO] 
731 [INFO] 
732 [INFO] 
733 [INFO] 
734 [INFO] 
735 [INFO] 
736 [INFO] 
737 [INFO] 
738 [INFO] 
739 [INFO] 
740 [INFO] 
741 [INFO] 
742 [INFO] 
743 [INFO] 
744 [INFO] 
745 [INFO] 
746 [INFO] 
747 [INFO] 
748 [INFO] 
749 [INFO] 
750 [INFO] 
751 [INFO] 
752 [INFO] 
753 [INFO] 
754 [INFO] 
755 [INFO] 
756 [INFO] 
757 [INFO] 
758 [INFO] 
759 [INFO] 
760 [INFO] 
761 [INFO] 
762 [INFO] 
763 [INFO] 
764 [INFO] 
765 [INFO] 
766 [INFO] 
767 [INFO] 
768 [INFO] 
769 [INFO] 
770 [INFO] 
771 [INFO] 
772 [INFO] 
773 [INFO] 
774 [INFO] 
775 [INFO] 
776 [INFO] 
777 [INFO] 
778 [INFO] 
779 [INFO] 
780 [INFO] 
781 [INFO] 
782 [INFO] 
783 [INFO] 
784 [INFO] 
785 [INFO] 
786 [INFO] 
787 [INFO] 
788 [INFO] 
789 [INFO] 
790 [INFO] 
791 [INFO] 
792 [INFO] 
793 [INFO] 
794 [INFO] 
795 [INFO] 
796 [INFO] 
797 [INFO] 
798 [INFO] 
799 [INFO] 
800 [INFO] 
801 [INFO] 
802 [INFO] 
803 [INFO] 
804 [INFO] 
805 [INFO] 
806 [INFO] 
807 [INFO] 
808 [INFO] 
809 [INFO] 
810 [INFO] 
811 [INFO] 
812 [INFO] 
813 [INFO] 
814 [INFO] 
815 [INFO] 
816 [INFO] 
817 [INFO] 
818 [INFO] 
819 [INFO] 
820 [INFO] 
821 [INFO] 
822 [INFO] 
823 [INFO] 
824 [INFO] 
825 [INFO] 
826 [INFO] 
827 [INFO] 
828 [INFO] 
829 [INFO] 
830 [INFO] 
831 [INFO] 
832 [INFO] 
833 [INFO] 
834 [INFO] 
835 [INFO] 
836 [INFO] 
837 [INFO] 
838 [INFO] 
839 [INFO] 
840 [INFO] 
841 [INFO] 
842 [INFO] 
843 [INFO] 
844 [INFO] 
845 [INFO] 
846 [INFO] 
847 [INFO] 
848 [INFO] 
849 [INFO] 
850 [INFO] 
851 [INFO] 
852 [INFO] 
853 [INFO] 
854 [INFO] 
855 [INFO] 
856 [INFO] 
857 [INFO] 
858 [INFO] 
859 [INFO] 
860 [INFO] 
861 [INFO] 
862 [INFO] 
863 [INFO] 
864 [INFO] 
865 [INFO] 
866 [INFO] 
867 [INFO] 
868 [INFO] 
869 [INFO] 
870 [INFO] 
871 [INFO] 
872 [INFO] 
873 [INFO] 
874 [INFO] 
875 [INFO] 
876 [INFO] 
877 [INFO] 
878 [INFO] 
879 [INFO] 
880 [INFO] 
881 [INFO] 
882 [INFO] 
883 [INFO] 
884 [INFO] 
885 [INFO] 
886 [INFO] 
887 [INFO] 
888 [INFO] 
889 [INFO] 
890 [INFO] 
891 [INFO] 
892 [INFO] 
893 [INFO] 
894 [INFO] 
895 [INFO] 
896 [INFO] 
897 [INFO] 
898 [INFO] 
899 [INFO] 
900 [INFO] 
901 [INFO] 
902 [INFO] 
903 [INFO] 
904 [INFO] 
905 [INFO] 
906 [INFO] 
907 [INFO] 
908 [INFO] 
909 [INFO] 
910 [INFO] 
911 [INFO] 
912 [INFO] 
913 [INFO] 
914 [INFO] 
915 [INFO] 
916 [INFO] 
917 [INFO] 
918 [INFO] 
919 [INFO] 
920 [INFO] 
921 [INFO] 
922 [INFO] 
923 [INFO] 
924 [INFO] 
925 [INFO] 
926 [INFO] 
927 [INFO] 
928 [INFO] 
929 [INFO] 
930 [INFO] 
931 [INFO] 
932 [INFO] 
933 [INFO] 
934 [INFO] 
935 [INFO] 
936 [INFO] 
937 [INFO] 
938 [INFO] 
939 [INFO] 
940 [INFO] 
941 [INFO] 
942 [INFO] 
943 [INFO] 
944 [INFO] 
945 [INFO] 
946 [INFO] 
947 [INFO] 
948 [INFO] 
949 [INFO] 
950 [INFO] 
951 [INFO] 
952 [INFO] 
953 [INFO] 
954 [INFO] 
955 [INFO] 
956 [INFO] 
957 [INFO] 
958 [INFO] 
959 [INFO] 
960 [INFO] 
961 [INFO] 
962 [INFO] 
963 [INFO] 
964 [INFO] 
965 [INFO] 
966 [INFO] 
967 [INFO] 
968 [INFO] 
969 [INFO] 
970 [INFO] 
971 [INFO] 
972 [INFO] 
973 [INFO] 
974 [INFO] 
975 [INFO] 
976 [INFO] 
977 [INFO] 
978 [INFO] 
979 [INFO] 
980 [INFO] 
981 [INFO] 
982 [INFO] 
983 [INFO] 
984 [INFO] 
985 [INFO] 
986 [INFO] 
987 [INFO] 
988 [INFO] 
989 [INFO] 
990 [INFO] 
991 [INFO] 
992 [INFO] 
993 [INFO] 
994 [INFO] 
995 [INFO] 
996 [INFO] 
997 [INFO] 
998 [INFO] 
999 [INFO] 
1000 [INFO] 
1001 [INFO] 
1002 [INFO] 
1003 [INFO] 
1004 [INFO] 
1005 [INFO] 
1006 [INFO] 
1007 [INFO] 
1008 [INFO] 
1009 [INFO] 
1010 [INFO] 
1011 [INFO] 
1012 [INFO] 
1013 [INFO] 
1014 [INFO] 
1015 [INFO] 
1016 [INFO] 
1017 [INFO] 
1018 [INFO] 
1019 [INFO] 
1020 [INFO] 
1021 [INFO] 
1022 [INFO] 
1023 [INFO] 
1024 [INFO] 
1025 [INFO] 
1026 [INFO] 
1027 [INFO] 
1028 [INFO] 
1029 [INFO] 
1030 [INFO] 
1031 [INFO] 
1032 [INFO] 
1033 [INFO] 
1034 [INFO] 
1035 [INFO] 
1036 [INFO] 
1037 [INFO] 
1038 [INFO] 
1039 [INFO] 
1040 [INFO] 
1041 [INFO] 
1042 [INFO] 
1043 [INFO] 
1044 [INFO] 
1045 [INFO] 
1046 [INFO] 
1047 [INFO] 
1048 [INFO] 
1049 [INFO] 
1050 [INFO] 
1051 [INFO] 
1052 [INFO] 
1053 [INFO] 
1054 [INFO] 
1055 [INFO] 
1056 [INFO] 
1057 [INFO] 
1058 [INFO] 
1059 [INFO] 
1060 [INFO] 
1061 [INFO] 
1062 [INFO] 
1063 [INFO] 
1064 [INFO] 
1065 [INFO] 
1066 [INFO] 
1067 [INFO] 
1068 [INFO] 
1069 [INFO] 
1070 [INFO] 
1071 [INFO] 
1072 [INFO] 
1073 [INFO] 
1074 [INFO] 
1075 [INFO] 
1076 [INFO] 
1077 [INFO] 
1078 [INFO] 
1079 [INFO] 
1080 [INFO] 
1081 [INFO] 
1082 [INFO] 
1083 [INFO] 
1084 [INFO] 
1085 [INFO] 
1086 [INFO] 
1087 [INFO] 
1088 [INFO] 
1089 [INFO] 
1090 [INFO] 
1091 [INFO] 
1092 [INFO] 
1093 [INFO] 
1094 [INFO] 
1095 [INFO] 
1096 [INFO] 
1097 [INFO] 
1098 [INFO] 
1099 [INFO] 
1100 [INFO] 
1101 [INFO] 
1102 [INFO] 
1103 [INFO] 
1104 [INFO] 
1105 [INFO] 
1106 [INFO] 
1107 [INFO] 
1108 [INFO] 
1109 [INFO] 
1110 [INFO] 
1111 [INFO] 
1112 [INFO] 
1113 [INFO] 
1114 [INFO] 
1115 [INFO] 
1116 [INFO] 
1117 [INFO] 
1118 [INFO] 
1119 [INFO] 
1120 [INFO] 
1121 [INFO] 
1122 [INFO] 
1123 [INFO] 
1124 [INFO] 
1125 [INFO] 
1126 [INFO] 
1127 [INFO] 
1128 [INFO] 
1129 [INFO] 
1130 [INFO] 
1131 [INFO] 
1132 [INFO] 
1133 [INFO] 
1134 [INFO] 
1135 [INFO] 
1136 [INFO] 
1137 [INFO] 
1138 [INFO] 
1139 [INFO] 
1140 [INFO] 
1141 [INFO] 
1142 [INFO] 
1143 [INFO] 
1144 [INFO] 
1145 [INFO] 
1146 [INFO] 
1147 [INFO] 
1148 [INFO] 
1149 [INFO] 
1150 [INFO] 
1151 [INFO] 
1152 [INFO] 
1153 [INFO] 
1154 [INFO] 
1155 [INFO] 
1156 [INFO] 
1157 [INFO] 
1158 [INFO] 
1159 [INFO] 
1160 [INFO] 
1161 [INFO] 
1162 [INFO] 
1163 [INFO] 
1164 [INFO] 
1165 [INFO] 
1166 [INFO] 
1167 [INFO] 
1168 [INFO] 
1169 [INFO] 
1170 [INFO] 
1171 [INFO] 
1172 [INFO] 
1173 [INFO] 
1174 [INFO] 
1175 [INFO] 
1176 [INFO] 
1177 [INFO] 
1178 [INFO] 
1179 [INFO] 
1180 [INFO] 
1181 [INFO] 
1182 [INFO] 
1183 [INFO] 
1184 [INFO] 
1185 [INFO] 
1186 [INFO] 
1187 [INFO] 
1188 [INFO] 
1189 [INFO] 
1190 [INFO] 
1191 [INFO] 
1192 [INFO] 
1193 [INFO] 
1194 [INFO] 
1195 [INFO] 
1196 [INFO] 
1197 [INFO] 
1198 [INFO] 
1199 [INFO] 
1200 [INFO] 
1201 [INFO] 
1202 [INFO] 
1203 [INFO] 
1204 [INFO] 
1205 [INFO] 
1206 [INFO] 
1207 [INFO] 
1208 [INFO] 
1209 [INFO] 
1210 [INFO] 
1211 [INFO] 
1212 [INFO] 
1213 [INFO] 
1214 [INFO] 
1215 [INFO] 
1216 [INFO] 
1217 [INFO] 
1218 [INFO] 
1219 [INFO] 
1220 [INFO] 
1221 [INFO] 
1222 [INFO] 
1223 [INFO] 
1224 [INFO] 
1225 [INFO] 
1226 [INFO] 
1227 [INFO] 
1228 [INFO] 
1229 [INFO] 
1230 [INFO] 
1231 [INFO] 
1232 [INFO] 
1233 [INFO] 
1234 [INFO] 
1235 [INFO] 
1236 [INFO] 
1237 [INFO] 
1238 [INFO] 
1239 [INFO] 
1240 [INFO] 
1241 [INFO] 
1242 [INFO] 
1243 [INFO] 
1244 [INFO] 
1245 [INFO] 
1246 [INFO] 
1247 [INFO] 
1248 [INFO] 
1249 [INFO] 
1250 [INFO] 
1251 [INFO] 
1252 [INFO] 
1253 [INFO] 
1254 [INFO] 
1255 [INFO] 
1256 [INFO] 
1257 [INFO] 
1258 [INFO] 
1259 [INFO] 
1260 [INFO] 
1261 [INFO] 
1262 [INFO] 
1263 [INFO] 
1264 [INFO] 
1265 [INFO] 
1266 [INFO] 
1267 [INFO] 
1268 [INFO] 
1269 [INFO] 
1270 [INFO] 
1271 [INFO] 
1272 [INFO] 
1273 [INFO] 
1274 [INFO] 
1275 [INFO] 
1276 [INFO] 
1277 [INFO] 
1278 [INFO] 
1279 [INFO] 
1280 [INFO] 
1281 [INFO] 
1282 [INFO] 
1283 [INFO] 
1284 [INFO] 
1285 [INFO] 
1286 [INFO] 
1287 [INFO] 
1288 [INFO] 
1289 [INFO] 
1290 [INFO] 
1291 [INFO] 
1292 [INFO] 
1293 [INFO] 
1294 [INFO] 
1295 [INFO] 
1296 [INFO] 
1297 [INFO] 
1298 [INFO] 
1299 [INFO] 
1300 [INFO] 
1301 [INFO] 
1302 [INFO] 
1303 [INFO] 
1304 [INFO] 
1305 [INFO] 
1306 [INFO] 
1307 [INFO] 
1308 [INFO] 
1309 [INFO] 
1310 [INFO] 
1311 [INFO] 
1312 [INFO] 
1313 [INFO] 
1314 [INFO] 
1315 [INFO] 
1316 [INFO] 
1317 [INFO] 
1318 [INFO] 
1319 [INFO] 
1320 [INFO] 
1321 [INFO] 
1322 [INFO] 
1323 [INFO] 
1324 [INFO] 
1325 [INFO] 
1326 [INFO] 
1327 [INFO] 
1328 [INFO] 
1329 [INFO] 
1330 [INFO] 
1331 [INFO] 
1332 [INFO] 
1333 [INFO] 
1334 [INFO] 
1335 [INFO] 
1336 [INFO] 
1337 [INFO] 
1338 [INFO] 
1339 [INFO] 
1340 [INFO] 
1341 [INFO] 
1342 [INFO] 
1343 [INFO] 
1344 [INFO] 
1345 [INFO] 
1346 [INFO] 
1347 [INFO] 
1348 [INFO] 
1349 [INFO] 
1350 [INFO] 
1351 [INFO] 
1352 [INFO] 
1353 [INFO] 
1354 [INFO] 
1355 [INFO] 
1356 [INFO] 
1357 [INFO] 
1358 [INFO] 
1359 [INFO] 
1360 [INFO] 
1361 [INFO] 
1362 [INFO] 
1363 [INFO] 
1364 [INFO] 
1365 [INFO] 
1366 [INFO] 
1367 [INFO] 
1368 [INFO] 
1369 [INFO] 
1370 [INFO] 
1371 [INFO] 
1372 [INFO] 
1373 [INFO] 
1374 [INFO] 
1375 [INFO] 
1376 [INFO] 
1377 [INFO] 
1378 [INFO] 
1379 [INFO] 
1380 [INFO] 
1381 [INFO] 
1382 [INFO] 
1383 [INFO] 
1384 [INFO] 
1385 [INFO] 
1386 [INFO] 
1387 [INFO] 
1388 [INFO] 
1389 [INFO] 
1390 [INFO]
```

B3: Then we just need to restart the namenode and check whether our jar file exist in our container

```
docker exec -it namenode bash
→ ~ docker restart namenode

namenode
→ ~ docker exec -it namenode bash
root@bbd0b037ecb0:/# ls /user/hadoop/target/
classes          hadoop-spark-project-1.0-SNAPSHOT.jar  maven-status
generated-sources  maven-archiver
root@bbd0b037ecb0:/# |
```

B4: Add the input file into input folder of project (persist which data/input/) and upload it into hdfs



```
docker exec -it namenode bash
root@bbd0b037ecb0:/data/input/lab2# ls
file_input1.txt  file_input2.txt
root@bbd0b037ecb0:/data/input/lab2#
```

```
docker exec -it namenode bash
root@bbd0b037ecb0:/data/input/lab2# hdfs dfs -put /data/input/lab2/ /user/hadoop/input/lab2_data/
put: '/user/hadoop/input/lab2_data/': No such file or directory: `hdfs://namenode:9000/user/hadoop/
/input/lab2_data'
root@bbd0b037ecb0:/data/input/lab2# hdfs dfs -mkdir /user/hadoop/input/lab2_data
root@bbd0b037ecb0:/data/input/lab2# hdfs dfs -put /data/input/lab2/ /user/hadoop/input/lab2_data/
2025-03-30 05:07:14,666 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTr
usted = false, remoteHostTrusted = false
2025-03-30 05:07:15,225 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTr
usted = false, remoteHostTrusted = false
root@bbd0b037ecb0:/data/input/lab2# |
```

B5: Run 3 test case: Normal, casesensitive, non-casensitive

```
root@bbd0b037ecb0:/data/input/lab2# hadoop jar /user/hadoop/target/hadoop-spark-project-1.0-SNAPSHOT.jar org.example.wordcountv2.WordCount2 /user/hadoop/input/lab2_data/lab2/ /user/hadoop/output/lab2_output/wordcount
2025-03-30 05:13:27,847 INFO client.RMProxy: Connecting to ResourceManager at resourcemanager/172.23.0.6:8032
2025-03-30 05:13:27,943 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.23.0.4:10200
2025-03-30 05:13:28,067 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1743311384508_0004
2025-03-30 05:13:28,130 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:13:28,203 INFO input.FileInputFormat: Total input files to process : 2
2025-03-30 05:13:28,228 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:13:28,245 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:13:28,660 INFO mapreduce.JobSubmitter: number of splits:2
2025-03-30 05:13:28,742 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:13:28,750 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1743311384508_0004
2025-03-30 05:13:28,751 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-30 05:13:28,842 INFO conf.Configuration: resource-types.xml not found
2025-03-30 05:13:28,843 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-03-30 05:13:29,080 INFO impl.YarnClientImpl: Submitted application application_1743311384508_0004
```

```
CPU time spent (ms)=820
Physical memory (bytes) snapshot=836419584
Virtual memory (bytes) snapshot=18756861952
Total committed heap usage (bytes)=726663168
Peak Map Physical memory (bytes)=309227520
Peak Map Virtual memory (bytes)=5139722240
Peak Reduce Physical memory (bytes)=220180480
Peak Reduce Virtual memory (bytes)=8482344960
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=55
File Output Format Counters
Bytes Written=67
org.example.wordcountv2.WordCount2$TokenizerMapper$CountersEnum
INPUT_WORDS=9
root@bbd0b037ecb0:/data/input/lab2#
```

```
root@bbd0b037ecb0:/data/input/lab2# hadoop jar /user/hadoop/target/hadoop-spark-project-1.0-SNAPSHOT.jar org.example.wordcountv2.WordCount2
root@bbd0b037ecb0:/data/input/lab2# hadoop jar /user/hadoop/target/hadoop-spark-project-1.0-SNAPSHOT.jar org.example.wordcountv2.WordCount2 -Dwordcount.case.sensitive=true /user/hadoop/input/lab2_data/lab2/ /user/hadoop/output/lab2_output/wordcount_casesen -skip /user/hadoop/heiper/skip_pattern.txt
2025-03-30 05:23:07,220 INFO client.RMProxy: Connecting to ResourceManager at resourcemanager/172.23.0.6:8032
2025-03-30 05:23:07,321 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.23.0.4:10200
2025-03-30 05:23:07,451 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1743311384508_0005
2025-03-30 05:23:07,513 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:23:07,591 INFO input.FileInputFormat: Total input files to process : 2
2025-03-30 05:23:07,615 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:23:07,632 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:23:07,640 INFO mapreduce.JobSubmitter: number of splits:2
2025-03-30 05:23:07,728 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:23:07,741 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1743311384508_0005
2025-03-30 05:23:07,742 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-30 05:23:07,865 INFO conf.Configuration: resource-types.xml not found
2025-03-30 05:23:07,865 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-03-30 05:23:08,107 INFO impl.YarnClientImpl: Submitted application application_1743311384508_0005
2025-03-30 05:23:08,145 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8088/proxy/application_1743311384508_0005/
```

```
root@bbd0b037cb0:/data/input/lab2# docker exec -it namenode bash
2025-03-30 05:23:07,865 INFO conf.Configuration: resource-types.xml not found
2025-03-30 05:23:07,865 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-03-30 05:23:08,107 INFO impl.YarnClientImpl: Submitted application application_1743311384508_0005
2025-03-30 05:23:08,145 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8088/proxy/application_1743311384508_0005/
2025-03-30 05:23:08,146 INFO mapreduce.Job: Running job: job_1743311384508_0005
2025-03-30 05:23:12,194 INFO mapreduce.Job: Job job_1743311384508_0005 running in uber mode : false
2025-03-30 05:23:12,195 INFO mapreduce.Job: map 0% reduce 0%
2025-03-30 05:23:16,231 INFO mapreduce.Job: map 50% reduce 0%
2025-03-30 05:23:17,245 INFO mapreduce.Job: map 100% reduce 0%
2025-03-30 05:23:20,269 INFO mapreduce.Job: map 100% reduce 100%
2025-03-30 05:23:21,280 INFO mapreduce.Job: Job job_1743311384508_0005 completed successfully
2025-03-30 05:23:21,334 INFO mapreduce.Job: Counters: 55
      File System Counters
          FILE: Number of bytes read=81
          FILE: Number of bytes written=692056
          FILE: Number of read operations=0
          FILE: Number of large read operations=0
          FILE: Number of write operations=0
          HDFS: Number of bytes read=323
          HDFS: Number of bytes written=55
          HDFS: Number of read operations=11
          HDFS: Number of large read operations=0
```

```
root@bbd0b037cb0:/data/input/lab2# docker exec -it namenode bash
      Bytes Read=55
      File Output Format Counters
          Bytes Written=55
      org.example.wordcountv2.WordCount$TokenizerMapper$CountersEnum
          INPUT_WORDS=9
root@bbd0b037cb0:/data/input/lab2# hadoop jar /user/hadoop/target/hadoop-spark-project-1.0-SNAPSHOT.jar org.example.wordcountv2.WordCount
-Dwordcount.case.sensitive=false /user/hadoop/input/lab2_data/lab2/ /user/hadoop/output/lab2_output/wordcount_casesen_false -skip /user/hadoop/helper/skip_pattern.txt
2025-03-30 05:25:16,566 INFO client.RMProxy: Connecting to ResourceManager at resourcemanager/172.23.0.6:8032
2025-03-30 05:25:16,767 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.23.0.4:10200
2025-03-30 05:25:16,992 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1743311384508_0006
2025-03-30 05:25:17,105 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:25:17,250 INFO input.FileInputFormat: Total input files to process : 2
2025-03-30 05:25:17,289 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:25:17,315 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:25:17,324 INFO mapreduce.JobSubmitter: number of splits:2
2025-03-30 05:25:17,447 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-03-30 05:25:17,459 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1743311384508_0006
2025-03-30 05:25:17,459 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-30 05:25:17,657 INFO conf.Configuration: resource-types.xml not found
2025-03-30 05:25:17,658 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
```

```
root@bbd0b037cb0:/data/input/lab2# docker exec -it namenode bash
2025-03-30 05:25:17,658 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-03-30 05:25:17,927 INFO impl.YarnClientImpl: Submitted application application_1743311384508_0006
2025-03-30 05:25:17,974 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8088/proxy/application_1743311384508_0006/
2025-03-30 05:25:17,975 INFO mapreduce.Job: Running job: job_1743311384508_0006
2025-03-30 05:25:24,075 INFO mapreduce.Job: Job job_1743311384508_0006 running in uber mode : false
2025-03-30 05:25:24,076 INFO mapreduce.Job: map 0% reduce 0%
2025-03-30 05:25:30,165 INFO mapreduce.Job: map 100% reduce 0%
2025-03-30 05:25:34,191 INFO mapreduce.Job: map 100% reduce 100%
2025-03-30 05:25:34,197 INFO mapreduce.Job: Job job_1743311384508_0006 completed successfully
2025-03-30 05:25:34,278 INFO mapreduce.Job: Counters: 55
      File System Counters
          FILE: Number of bytes read=77
          FILE: Number of bytes written=692071
          FILE: Number of read operations=0
          FILE: Number of large read operations=0
          FILE: Number of write operations=0
          HDFS: Number of bytes read=323
          HDFS: Number of bytes written=46
          HDFS: Number of read operations=11
          HDFS: Number of large read operations=0
          HDFS: Number of write operations=2
          HDFS: Number of bytes read erasure-coded=0
```

B6: Cat them and tee into our data/output which persist which the output folder of project

```
root@bbd0b037ecb0:/data/input/lab2# hdfs dfs -cat /user/hadoop/output/lab2_output/wordcount/part-r-00000 | tee /data/output/wordcount2.txt
2025-03-30 05:15:23,488 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Bye 1
Goodbye 1
Hadoop 1
Hello 2
World! 1
World, 1
hadoop. 1
to 1
root@bbd0b037ecb0:/data/input/lab2#
```

```
root@bbd0b037ecb0:/data/input/lab2# hdfs dfs -cat /user/hadoop/output/lab2_output/wordcount_casesen/part-r-00000 | tee /data/output/wordcount2_sen.txt
2025-03-30 05:28:16,448 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Bye 1
Goodbye 1
Hadoop 1
Hello 2
World 2
hadoop 1
to 1
root@bbd0b037ecb0:/data/input/lab2# hdfs dfs -cat /user/hadoop/output/lab2_output/wordcount_casesen_false/part-r-00000 | tee /data/output/wordcount2_sen_false.txt
2025-03-30 05:28:29,988 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
bye 1
goodbye 1
hadoop 2
hello 2
to 1
world 2
root@bbd0b037ecb0:/data/input/lab2#
```

B7: See the result

The screenshot shows the IntelliJ IDEA interface with the project structure on the left and three code editor panes on the right.

- Project Tree:** The project is named "hadoop-spark-project". It contains several modules and files:
 - lab1:** Contains `lab1_data`, `lab1`, and `lab1_output`.
 - lab2:** Contains `file_input1.txt`, `file_input2.txt`, and `skip_pattern.txt`.
 - lab:** Contains `lab2`.
 - output:** Contains `airsensor.txt`, `electricity.txt`, `number.txt`, `sales.txt`, `weblog.txt`, `websitetime.txt`, `wordcount.txt`, `wordcount2.txt`, `wordcount2_sen.txt`, and `wordcount2_sen_false.txt`.
 - src/main/java:** Contains packages `org.example` with classes `airsensor`, `electricity`, `employee`, `number`, `polygon`, `sale`, `weblog`, `websitetime`, `wordcount`, and `WordCount2`.
- Code Editors:** Three files are open:
 - `wordcount2.txt`: Shows the original input data with mixed case and punctuation.
 - `wordcount2_sen.txt`: Shows the output where all words are converted to lowercase.
 - `wordcount2_sen_false.txt`: Shows the output where punctuation marks like '!' and ',' are preserved.

Câu 3:

Kết quả câu a,b:

1 Count A: 28, Count AB: 1, P(2810 11076)	0.03571
2 Count A: 55, Count AB: 1, P(10045 2583)	0.01818
3 Count A: 30, Count AB: 1, P(1404 14056)	0.03333
4 Count A: 26, Count AB: 1, P(1048 6509)	0.03846
5 Count A: 30, Count AB: 1, P(1409 14056)	0.03333
6 Count A: 426, Count AB: 1, P(6956 2523)	0.00235
7 Count A: 92, Count AB: 2, P(2049 5982)	0.02174
8 Count A: 107, Count AB: 1, P(1814 9823)	0.00935
9 Count A: 17, Count AB: 1, P(5948 9078)	0.05882
10 Count A: 105, Count AB: 1, P(11588 1073)	0.00952
11 Count A: 155, Count AB: 1, P(1052 1246)	0.00645
12 Count A: 92, Count AB: 1, P(2052 5982)	0.01087
13 Count A: 96, Count AB: 1, P(6845 3579)	0.01042
14 Count A: 18, Count AB: 1, P(7910 2981)	0.05556
15 Count A: 3, Count AB: 1, P(604 11929) 0.33333	
16 Count A: 92, Count AB: 1, P(2059 5982)	0.01087
17 Count A: 91, Count AB: 1, P(2004 6353)	0.01099
18 Count A: 143, Count AB: 1, P(2004 1090)	0.00699
19 Count A: 143, Count AB: 1, P(2005 1090)	0.00699
20 Count A: 13, Count AB: 1, P(6813 12200)	0.07692
21 Count A: 155, Count AB: 1, P(1063 1246)	0.00645
22 Count A: 14, Count AB: 1, P(1415 183) 0.07143	
23 Count A: 155, Count AB: 1, P(1066 1246)	0.00645
24 Count A: 155, Count AB: 3, P(1067 1246)	0.01935
25 Count A: 155, Count AB: 1, P(1068 1246)	0.00645
26 Count A: 426, Count AB: 2, P(6979 2523)	0.00469
27 Count A: 20, Count AB: 3, P(989 2467) 0.15000	
28 Count A: 91, Count AB: 1, P(2014 6353)	0.01099
29 Count A: 91, Count AB: 1, P(2015 6353)	0.01099
30 Count A: 26, Count AB: 1, P(1079 6509)	0.03846
31 Count A: 96, Count AB: 1, P(6864 3579)	0.01042
32 Count A: 107, Count AB: 1, P(10943 816)	0.00935
33 Count A: 18, Count AB: 1, P(7931 2981)	0.05556
34 Count A: 96, Count AB: 2, P(6869 3579)	0.02083
35 Count A: 92, Count AB: 1, P(2078 5982)	0.01087
36 Count A: 18, Count AB: 1, P(7934 2981)	0.05556
37 Count A: 18, Count AB: 1, P(7938 2981)	0.05556
38 Count A: 366, Count AB: 1, P(7009 734)	0.00273
39 Count A: 49, Count AB: 1, P(2612 4152)	0.02041
40 Count A: 107, Count AB: 3, P(1848 9823)	0.02804
41 Count A: 831, Count AB: 2, P(2202 389)	0.00241
42 Count A: 3, Count AB: 1, P(1772 9289) 0.33333	

Kết quả câu c,d:

1 Count A: 37, Count ABC: 1, P(998,878 3640)	0.02703
2 Count A: 58, Count ABC: 1, P(996,857 2306)	0.01724
3 Count A: 201, Count ABC: 1, P(997,8559 3638)	0.00498
4 Count A: 44, Count ABC: 1, P(999,9949 2305)	0.02273
5 Count A: 11, Count ABC: 1, P(986,94 3639)	0.09091
6 Count A: 25, Count ABC: 1, P(994,8483 2304)	0.04000
7 Count A: 19, Count ABC: 1, P(983,5646 2303)	0.05263
8 Count A: 3, Count ABC: 1, P(9527,3539 2302)	0.33333
9 Count A: 34, Count ABC: 1, P(9879,824 2301)	0.02941
10 Count A: 118, Count ABC: 1, P(997,965 2300)	0.00847
11 Count A: 53, Count ABC: 1, P(989,4304 3630)	0.01887
12 Count A: 7, Count ABC: 1, P(817,2968 3631)	0.14286
13 Count A: 45, Count ABC: 1, P(985,921 3632)	0.02222
14 Count A: 125, Count ABC: 1, P(9925,705 3633)	0.00800
15 Count A: 4, Count ABC: 1, P(9333,8985 3634)	0.25000
16 Count A: 34, Count ABC: 1, P(99,78 2309)	0.02941
17 Count A: 264, Count ABC: 1, P(99,78 3635)	0.00379
18 Count A: 27, Count ABC: 1, P(941,924 2308)	0.03704
19 Count A: 46, Count ABC: 1, P(9913,8679 3636)	0.02174
20 Count A: 4, Count ABC: 1, P(987,4836 2307)	0.25000
21 Count A: 7, Count ABC: 1, P(829,1603 3637)	0.14286
22 Count A: 23, Count ABC: 1, P(9682,812 3650)	0.04348
23 Count A: 60, Count ABC: 1, P(9985,7502 3651)	0.01667
24 Count A: 5, Count ABC: 1, P(9360,681 2320)	0.20000
25 Count A: 177, Count ABC: 1, P(9943,62 0)	0.00565
26 Count A: 33, Count ABC: 1, P(9974,661 2317)	0.03030
27 Count A: 11, Count ABC: 1, P(9551,9029 3649)	0.09091
28 Count A: 266, Count ABC: 1, P(9943,6358 1)	0.00376
29 Count A: 54, Count ABC: 1, P(992,956 2316)	0.01852
30 Count A: 549, Count ABC: 1, P(9986,16431 2)	0.00182
31 Count A: 96, Count ABC: 1, P(9964,897 2315)	0.01042
32 Count A: 64, Count ABC: 1, P(984,789 2314)	0.01563
33 Count A: 8, Count ABC: 1, P(9407,7205 3)	0.12500
34 Count A: 65, Count ABC: 1, P(9879,828 2313)	0.01538
35 Count A: 19, Count ABC: 1, P(986,829 4)	0.05263
36 Count A: 8, Count ABC: 1, P(911,5996 2312)	0.12500
37 Count A: 5, Count ABC: 1, P(6128,3735 2311)	0.20000
38 Count A: 132, Count ABC: 1, P(979,6649 2310)	0.00758
39 Count A: 69, Count ABC: 1, P(9937,2747 3641)	0.01449
40 Count A: 158, Count ABC: 1, P(989,5802 3642)	0.00633
41 Count A: 60, Count ABC: 1, P(9703,3877 3643)	0.01667
42 Count A: 111, Count ABC: 1, P(9706,10721 3644)	0.00901

Câu 4:

Kết quả câu a:

You, 17 minutes ago | 1 author (Yours)

1	2022-02-23	8
2	2022-02-25	8
3	2022-03-02	8
4	2022-03-11	7
5	2022-03-15	7
6	2022-03-17	7
7	2022-03-22	7
8	2022-03-24	7
9	2022-04-01	7
10	2022-04-05	7
11	2022-04-07	7
12	2022-04-12	7
13	Total	7
14	Total	7
15	-	

Kết quả câu b:

	You, 17 minutes ago 1 author (You)
1	Bia Heineken lon (thùng 24 lon) 408695.65
2	Bia Tiger lon (thùng 24 lon) 344347.83
3	Bia lon 333 Sài gòn (thùng 24 lon) 253043.48
4	Bánh chưng loại 1,5kg/cái 176086.96
5	Bánh hộp Danisa butter cookie 454gr 138695.65
6	Bí xanh 31043.48
7	Bưởi nǎm roi (da xanh) (trái 1 kg) 59347.83
8	Bưởi nǎm roi (trái 1kg) 44347.83
9	Bắp cải Đà Lạt (lặt sạch) 29130.43
10	Bột ngọt Ajinomoto (VN 453gr/gói) 32000.00
11	Cam sành (loại 4-5 trái/kg) 45869.57
12	Chân giò 168695.65
13	Chả Bò 220000.00
14	Chả lụa 214782.61
15	Cà chua 36608.70
16	Cà rốt Đà Lạt 31565.22
17	Cá basa 60000.00
18	Cá bạc má 80000.00
19	Cá diêu hồng (con trên 0,7kg) 70434.78
20	Cá hường 90000.00
21	Cá lóc nuôi bè (0,5kg/con) 79565.22
22	Cá thu 276086.96
23	Cá thác lác 350000.00
24	Cải xanh 59869.57
25	Dâu ăn tôtng hợp Nakydaco (chai 1 lít) 40000.00
26	Dâu ăn tôtng hợp Neptune (chai 1 lit) 55000.00
27	Dâu ăn tôtng hợp Tường An (chai 1 lít) 45000.00
28	Dưa hấu tròn (trái 4-5kg) 21521.74
29	Dưa leo 27652.17
30	Gạo Tám thơm 20000.00
31	Gà công nghiệp làm săn 73260.87
32	Gà mái ta làm săn (con 1-1,5kg) 138260.87
33	Gà tam hoàng làm săn (con 1-1,5kg) 87391.30
34	Gạo nàng thơm chợ Đào 24782.61
35	Gạo trắng Jasmine 17391.30
36	Gạo tẻ trắng IR 64 25% tám 14608.70
37	Hoa huệ 118695.65
38	Hành tím Gò Công 45000.00 You, 17 minutes ago • [UPDA]: Qu
39	Hạt dưa 160000.00
40	Khoai tây Đà Lạt 30434.78
41	Khô'qua 28521.74
42	Lê TQ 35000.00

Kết quả câu c:

You, 18 minutes ago | 1 author (You)

- 1 Bia Heineken lon (thùng 24 lon) Min: 400000.00, Max: 410000.00
- 2 Bia Tiger lon (thùng 24 lon) Min: 340000.00, Max: 345000.00
- 3 Bia lon 333 Sài gòn (thùng 24 lon) Min: 240000.00, Max: 255000.0
- 4 Bánh chưng loại 1,5kg/cái Min: 150000.00, Max: 180000.00
- 5 Bánh hộp Danisa butter cookie 454gr Min: 130000.00, Max: 140000.00
- 6 Bí xanh Min: 18000.00, Max: 40000.00
- 7 Bưởi nǎm roi (da xanh) (trái 1 kg) Min: 55000.00, Max: 60000.00
- 8 Bưởi nǎm roi (trái 1kg) Min: 40000.00, Max: 45000.00
- 9 Bắp cải Đà Lạt (lặt sạch) Min: 20000.00, Max: 35000.00
- 10 Bột ngọt Ajinomoto (VN 453gr/gói) Min: 32000.00, Max: 32000.00
- 11 Cam sành (loại 4-5 trái/kg) Min: 40000.00, Max: 70000.00
- 12 Chân giò Min: 160000.00, Max: 170000.00
- 13 Chả Bò Min: 220000.00, Max: 220000.00
- 14 Chả lụa Min: 180000.00, Max: 220000.00
- 15 Cà chua Min: 25000.00, Max: 55000.00
- 16 Cà rốt Đà Lạt Min: 28000.00, Max: 40000.00
- 17 Cá basa Min: 60000.00, Max: 60000.00
- 18 Cá bạc má Min: 80000.00, Max: 80000.00
- 19 Cá diêu hồng (con trên 0,7k/g) Min: 65000.00, Max: 75000.00
- 20 Cá hường Min: 90000.00, Max: 90000.00
- 21 Cá lóc nuôi bè (0,5kg/con) Min: 70000.00, Max: 90000.00
- 22 Cá thu Min: 250000.00, Max: 350000.00
- 23 Cá thác lác Min: 350000.00, Max: 350000.00
- 24 Cải xanh Min: 20000.00, Max: 300000.00
- 25 Dâu ăn tông hợp Nakydaco (chai 1 lít) Min: 40000.00, Max: 40000
- 26 Dâu ăn tông hợp Neptune (chai 1 lit) Min: 55000.00, Max: 55000.00
- 27 Dâu ăn tông hợp Tường An (chai 1 lít) Min: 45000.00, Max: 45000.
- 28 Dưa hấu tròn (trái 4-5kg) Min: 20000.00, Max: 25000.00
- 29 Dưa leo Min: 20000.00, Max: 30000.00
- 30 Gạo Tám thơm Min: 20000.00, Max: 20000.00
- 31 Gà công nghiệp làm săn Min: 50000.00, Max: 80000.00
- 32 Gà mái ta làm săn (con 1-1,5kg) Min: 120000.00, Max: 220000.00
- 33 Gà tam hoàng làm săn (con 1-1,5kg) Min: 70000.00, Max: 90000.00
- 34 Gạo nàng thơm chợ Đào Min: 23000.00, Max: 26000.00
- 35 Gạo trắng Jasmine Min: 15000.00, Max: 28000.00
- 36 Gạo tẻ trắng IR 64 25% tẩm Min: 12000.00, Max: 15000.00
- 37 Hoa huệ Min: 110000.00, Max: 120000.00
- 38 Hành tím Gò Công Min: 45000.00, Max: 45000.00
- 39 Hạt dưa Min: 160000.00, Max: 160000.00
- 40 Khoai tây Đà Lạt Min: 30000.00, Max: 35000.00
- 41 Khô'qua Min: 20000.00, Max: 35000.00
- 42 Lê TQ Min: 35000.00, Max: 35000.00